



Community Webs:

New Developments in Access and Use

October 25, 2022

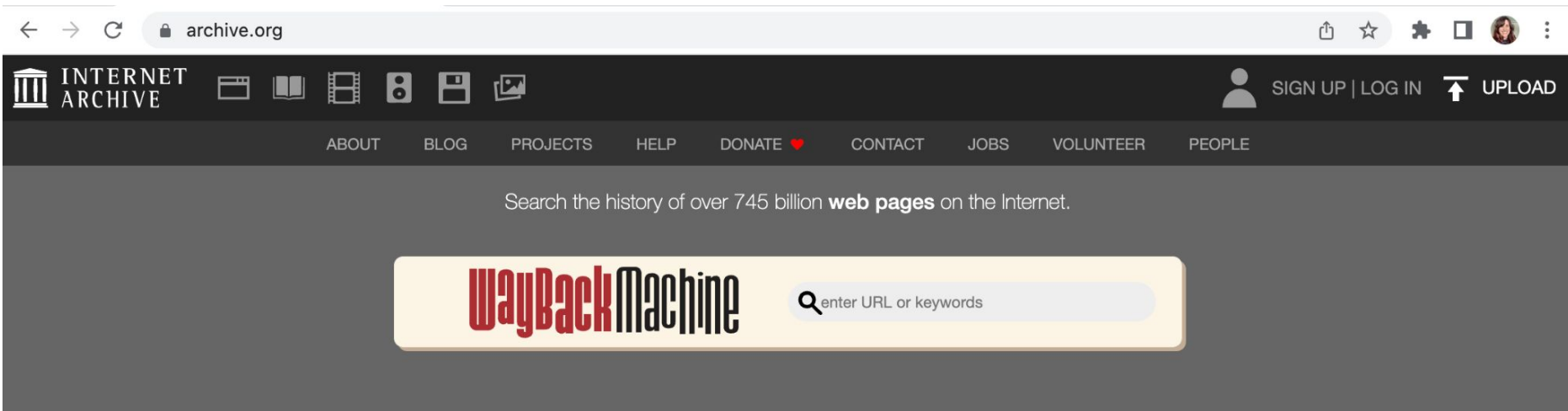


Overview of Web Archive Access



Web Archive Access - Wayback Machine (archive.org)

- View historic captures of websites (625 billion pages)
- Search by URL or keyword (not full text search)
- Includes Internet Archive-run crawling as well as Archive-It and other partner-directed crawls





Web Archive Access - archive-it.org

- Access to all public collections in Archive-It
- Browse by URL and search metadata and full text
- Browse by organization, collection or sites

The screenshot displays the Archive-It website interface. At the top, there is a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. A login button is located in the top right corner. Below the navigation bar, a banner reads "Welcome to Archive-It!" and "Contact Us to learn more about our products and services." The main content area features three featured collections:

- Amateur Mormon Historian**
By Brigham Young University
Features the lifestyle and culture of Mormons through self-published blogs.
- Ukraine Conflict**
By Internet Archive Global Events
This collection seeks to document conflict in Ukraine as it progresses. Content includes news outlets, social media, blogs, and government websites. Sites are written in English, Russian,...
- Everglades Explorer -- EAPRA (Assorted PDF & Report Archive)**
By Florida International University Libraries
An archive of digital government and non-government organization (NGO) documents and reports, representing the Greater Everglades watershed and adjacent ecosystems, including...

At the bottom, there is a section for "Explore Collecting Organizations" with a search bar and a link to "Show All Organizations".



Web Archive Access - communitywebs.archive-it.org

- Access to all public Community Webs member collections
- Browse by URL and search metadata and full text
- Featured Collection topics allow access to similar collections across members.

The screenshot shows the Community Webs homepage. At the top is a navigation bar with links: Home, About, Collections, Participants, Curriculum, News, and Apply. Below the navigation bar is a large banner image with the text "Community Webs" and "Empowering Cultural Heritage Organizations to Create Community History Web Archives". A search bar is present in the banner with the text "Search across 100s of Web Collections" and a "Search" button. Below the banner, there are two main sections: "Explore Organizations by Location" on the left, which features a world map with blue location pins, and "About this Program" on the right, which contains a paragraph about the program's mission and a link "More about this program". At the bottom, there is a "Featured Collections" section with icons for a globe and a book.

The screenshot shows the search results page for the query "#DaytonStrong". The page has a navigation bar with links: Home, About, Collections, Participants, Curriculum, News, and Apply. Below the navigation bar is a search bar with the text "Search" and a "Clear" button. The search results are displayed in a two-column layout. The left column is titled "Narrow Your Results" and lists organizations: Cleveland Public Library (39), Grand Rapids Public Library (36), East Baton Rouge Parish Library (32), New Brunswick Free Public Library (24), Birmingham Public Library (21), DC Public Library (19), Henderson County Libraries (14), Pueblo City-County Library District (14), State Library of Pennsylvania (12), and Choctaw Cultural Center (11). The right column shows the search results for "#DaytonStrong". It includes a "Previous Page" button, "Page 1 of 33 (653 Total Results)", and a "Next Page" button. The results for "#DaytonStrong" include the organization "Dayton Metro Library", a description of the collection, the creator "Dayton Metro Library", the publisher "Dayton Metro Library", the rights "http://rightsstatements.org/vocab/InC-EDU/1.0/", and the type "Archived Website". Below this, there is a section for "#Syllabus" with the organization "Schomburg Center for Research in Black Culture, New York Public Library", a description of the collection, and a link to the collection.



Community Webs Access Site Demo



Metadata Integration Updates

What is the Digital Public Library of America (DPLA)?

Digital Public Library of America: <https://dp.la/>

- Launched in 2013
- Metadata aggregation
- Content Hubs (including Community Webs)





What Community Webs content will be included in DPLA?

- US-based Community Webs members
- Public Collections and their seeds
 - Must have at least Collection level rights statement
- Member-created metadata, plus automated metadata as needed
- Pointers back to the archived content
- Updates on a regular basis (approximately quarterly): Full replacement each time
- Launched late August 2022 with >4,800 items



Collection Level Metadata in DPLA

- Title (Collection Name)
- Rights - rightsstatements.org or creativecommons.org URL
- Link to resource: Archive-It collection page, eg:
<https://archive-it.org/collections/1506>
- Image:
 - Collection image or account image
- Description
- Subject(s)
- Coverage, Creator, Language, Relation, Date



Seed Level Metadata in DPLA

- Title
 - If no title exists, we grab title from live web html, otherwise use URL
- Link to resource: Seed calendar page, eg:
https://wayback.archive-it.org/8092/*/https://www.boisestate.edu/
- Rights
 - Inherits from collection level if not set for individual seeds



Seed Level Metadata in DPLA - continued

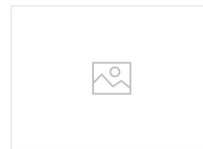
- Description
 - If description exists use that, if not, we try to scrape from live site
- Type
 - Interactive Resource (this is set by DPLA)
- Format
 - If no metadata is provided will use “Archived Website”
- Collection (links to collection level metadata)
- Coverage, Creator, Language, Relation, Date if provided by you!



Seed Level Metadata in DPLA - Image

Image: We try to get a representative image of the website by checking:

1. Does the site have a preferred image for Twitter?
2. Was the site crawled using Brozzler (which takes screenshots)?
3. Does the site have an easily identifiable logo?
4. Does the site have a favicon?



CPS to resume food distribution program after safety concerns led to day-long suspension - Chicago Sun-Times CPS to resume food distribution program a...

Nader Issa

The district, the nation's third largest, has given out more than 12.5 million meals since the start of the coronavirus pandemic.;This collection centers and documents the African diasporan experience...

[View Full Item](#) in Schomburg Center for Research in Black Culture



Home | Schomburg Education Education at the Schomburg Center for Research in Black

Schomburg Education Department, Schomburg Center for Research in Black Culture, The New York Public Library

Home | Schomburg Education;Collections of websites, resources, and recordings created by the Schomburg Center Education department.

[View Full Item](#) in Schomburg Center for Research in Black Culture



Using NYPL resources to enrich your African American History studies | The New York Public Library Using NYPL resources to enrich your African America...

Lynda Kennedy

Toussaint Louverture, about 1795 (NYPL Digital Gallery)Use the resources of NYPL to engage your students with the rich and complex history of African Americans.;Collection of blog posts written by Sch...

[View Full Item](#) in Schomburg Center for Research in Black Culture



Rights Statements in DPLA

- Must be expressed as a URL, rather than the name of the rights designation
 - For example: <http://rightsstatements.org/vocab/InC-EDU/1.0/>
rather than IN COPYRIGHT - EDUCATIONAL USE PERMITTED
- Must come from rightsstatements.org or creativecommons.org
 - <https://rightsstatements.org/page/1.0/?language=en>
 - <https://creativecommons.org/about/cclicenses/>
- Can include local rights statement as well but it cannot contradict the standard statement



DPLA Demo



DPLA - Remaining questions and points for consideration

- Default rights statement?
 - Copyright not evaluated:
<https://rightsstatements.org/page/CNE/1.0/?language=en>
- Organization name in Archive-It vs Organization name in other DPLA hub(s)

Contributing Institution	—
San Francisco Public Library	17,402
Partner	—
California Digital Library	16,482
Community Webs	920

Contributing Institution	—
Athens Regional Library System	134
Partner	—
Community Webs	134

Contributing Institution	—
Athens-Clarke County Library	1,640
Partner	—
Digital Library of Georgia	1,640



MetaData Integration: OurDigitalWorld



search.ourontario.ca

Apps Internet Archive -... My Drive - Google... Maps Mailchimp Dashbo... Community Webs

Reading List

Discover...
our stories
our history
our Ontario

AUDIO
COLLECTION
IMAGE
OBJECT
TEXT
VIDEO

ABOUT FAQ HELP | EN FRANÇAIS | SHARE

SEARCH

ADVANCED SEARCH

BROWSE BY CONTRIBUTOR

Discover over 2.3 million items and collections from libraries, archives, museums, historical societies, community groups, and government ministries.

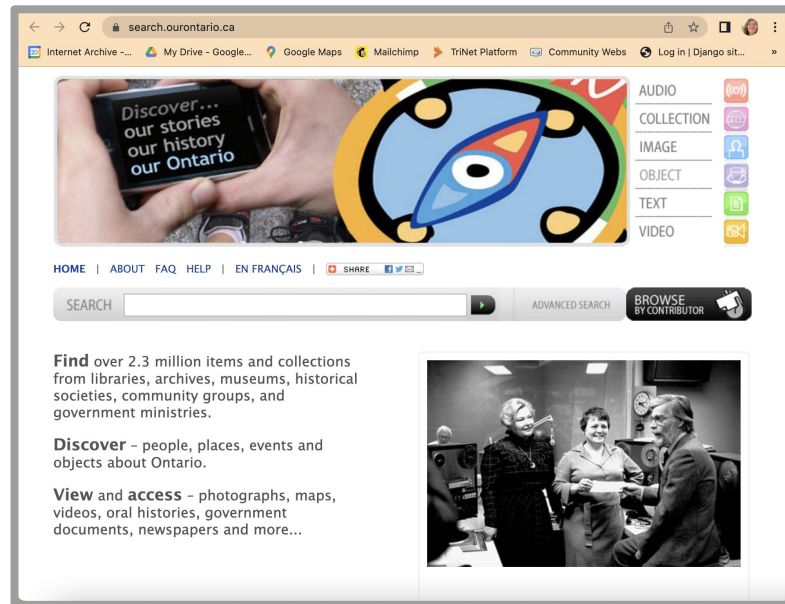
Discover – people, places, events and objects about Ontario.

View and access – photographs, maps, videos, oral histories, government documents, newspapers and more...



OurOntario.ca

- Search portal for multiple heritage organizations' collections
- ~300 heritage organizations are indexed
- Currently searches nearly 3 million items
- Ontario's largest online collection of newspapers
- Award-winning platform





OurDigitalWorld x Community Webs

- Ontario-based Community Webs content
- Opt-in participation
- Metadata to align with ODW standards
- Reducing web archive silos
- Determining representative image

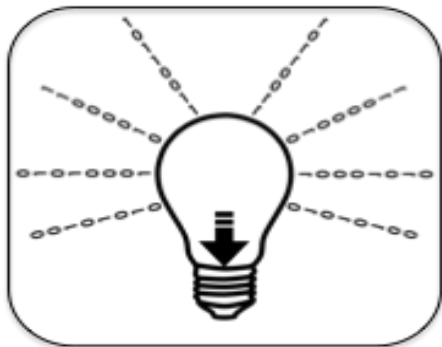




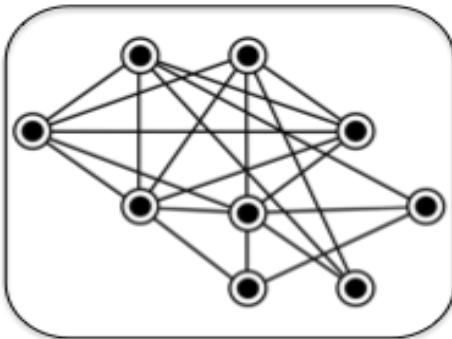
Use of Web Archive Data

Archive-It Research Services (2014)

Web Archive Datasets



WAT Datasets
(Web Archive
Transformation)
Key Metadata from Every
Resource



LGA Datasets
(Longitudinal
Graph Analysis)
What Links to What
over Time



WANE Datasets
(Web Archive
Named Entities)
Names of People, Places,
Organizations

Archives Unleashed Project (2017-2020)

- Est. 2017 with funding from the Andrew W. Mellon Foundation
- Recognizes the **critical role of web archives** for scholars studying the 1990s onward
- **Lower barriers to access & use of web archives**
- Project Priorities:
 - Tool Development
 - Resource Support
 - Community Engagement

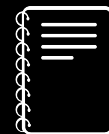
Tools



Toolkit



Cloud



Notebooks

Resources



Learning Guides

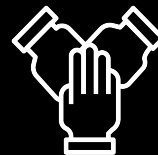


Video tutorials

Community



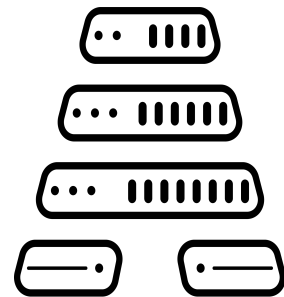
Datathon Series



Collaborations



Archives
Unleashed

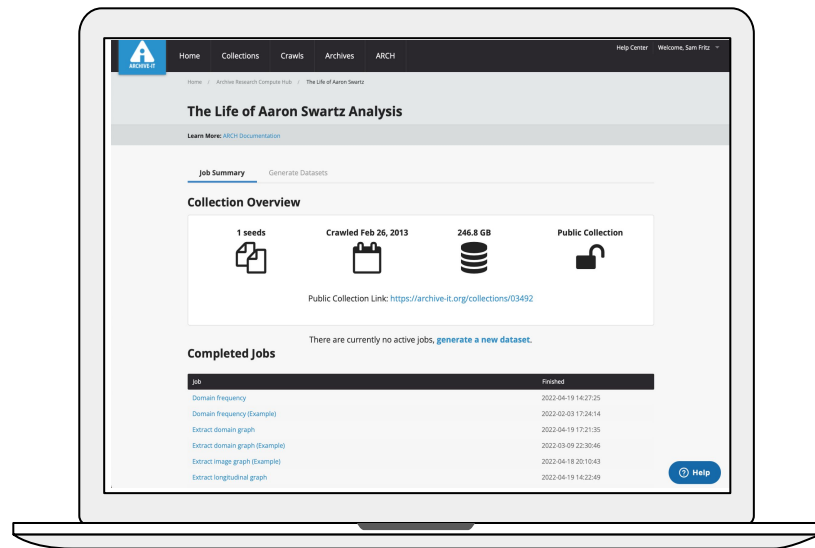


ARCH

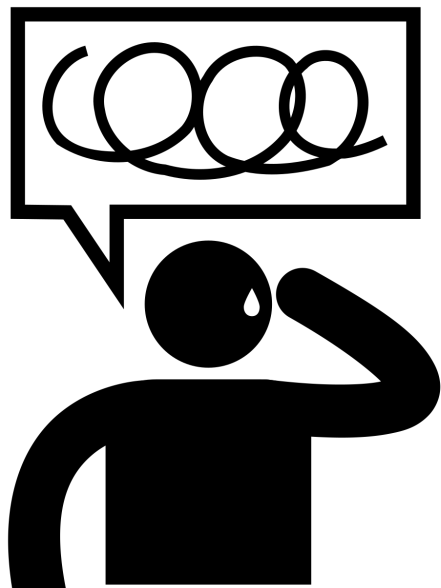


ARCH Features

- Interactive, familiar environment for current Archive-It subscribers
- Transforms Archive-It collections into research datasets for analysis
- Generate and download over a dozen datasets
- Standardized dataset format as CSVs
- In-browser visualizations and data previews that presents a glimpse into collection content



What can you do with CSV files?



- What is your research question?
 - ◆ Text/language based?
 - ◆ Network?
 - ◆ Other
- What tools/methods are you familiar with?
Are there any you're interested in further exploring?
- [ARCH User Documentation](#) provides some examples of using tools like Gephi, Voyant or methods like topic modelling or geoparsing with ARCH datasets



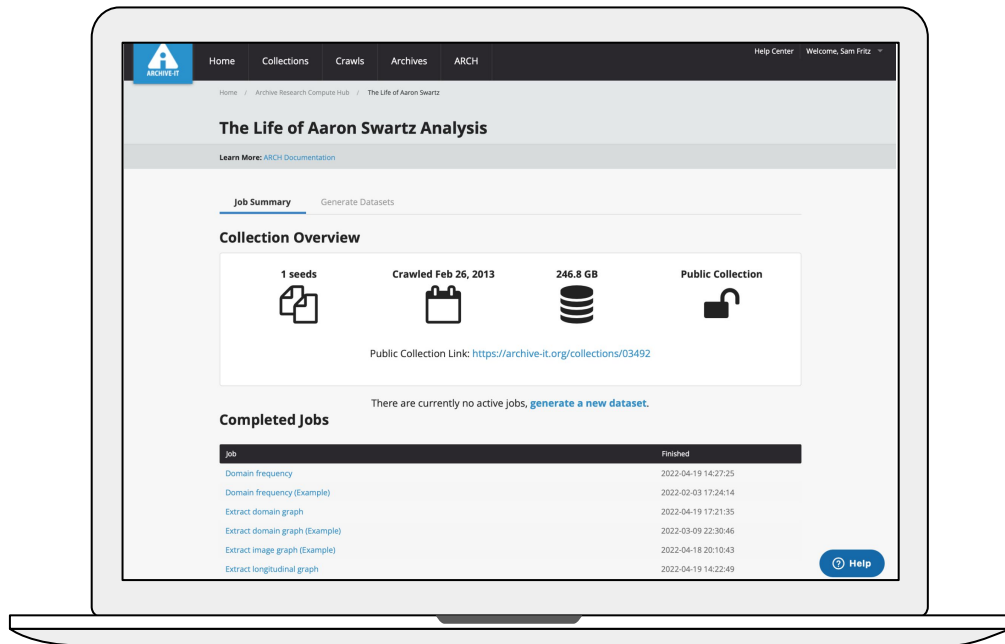
ARCH Demo



ARCH Pilot

Interested in working with ARCH in its pilot phase?

Fill out [this form](#) and the team will be in touch.





Thank you! And Resources

- [Community Webs Access Portal](#)
- Archive-It Resources:
 - [Metadata in Archive-It](#)
 - [Archive-It Integrations](#)
- DPLA Resources:
 - [Standardized Rights Statements](#)
- [OurDigitalWorld](#)
- [ARCH Documentation](#)